

University of Groningen

Biomarker discovery for cervical cancer

Govorukhina, Natalia I.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2007

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Govorukhina, N. I. (2007). *Biomarker discovery for cervical cancer*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter IV.

Influence of clotting time on the protein composition of serum samples based on LC-MS data

N.I. Govorukhina, M. de Vries, T.H. Reijmers, P. Horvatovich, A.G.J. van der Zee, and R. Bischoff

Manuscript in preparation

Abstract

Many large banks of serum samples were acquired prior to the widespread use of proteomics in biomarker research. An important parameter that is often not exactly known is clotting time. We therefore investigated the influence of clotting time on the protein and peptide composition of serum by label-free and stable-isotope labeling techniques. The label-free analysis of trypsin-digested serum showed that the overall pattern of LC-MS data is not affected by clotting times varying from 2 to 8h. However, univariate and multivariate statistical analyses revealed that proteins that are directly involved in blood clot formation, such as the clotting-derived fibrinopeptides, change significantly. This is most easily detected in the supernatant of acid-precipitated, immunodepleted serum. Stable-isotope labeling techniques show that truncated or phosphorylated forms of Fibrinopeptides A and B increase or decrease depending on clotting time. These patterns can be easily recognized and should be taken into consideration when using serum sample collections of which the clotting time is not known. Leucine-rich alpha-2-glycoprotein (P02750) was down-regulated in all samples compared to the sample for which the clotting time was one hour based on iTRAQ labeling data.

1. Introduction

The discovery and validation of biomarkers for early diagnosis of disease at a stage where successful therapy is still possible is an important goal of modern biomedical research. To achieve this goal, high-resolution analytical techniques are applied to complex clinical samples, mostly body fluids. Serum is a body fluid that is representative of the composition of soluble proteins and peptides in blood and is thus a suitable starting material for biomarker discovery studies. Moreover, many existing large sample collections at major hospitals consist of serum that is stored frozen at -80°C . Since these collections may well contain important information about the health status of the corresponding patients and controls, especially when they have been followed over time, it is critical to evaluate under which conditions it is possible to compare samples from these collections with modern proteomics approaches.

The generation of serum requires that blood be coagulated and that the cellular components as well as the blood clot be removed by centrifugation or filtration. It has notably been argued that the time and conditions under which blood is allowed to clot (clotting time) are important parameters that must be controlled and kept constant in order to compare protein and peptide profiles [1-5]. However, most existing sample collections have not been obtained with subsequent proteomics analyses in mind and clotting time has often not been rigorously controlled. Many of the studies evaluating the influence of pre-analytical parameters on serum protein composition have been performed by Surface-Enhanced Laser Desorption Ionization Time-Of-Flight Mass Spectrometry (SELDI-TOF-MS) [4], a method that suffers from rather poor concentration sensitivity and that may be prone to mass spectrometric artifacts [6]. More sensitive approaches using enrichment of proteins and peptides on magnetic bead separators or by Liquid Chromatography (LC) followed by Matrix-Assisted Laser Desorption Ionization Time-Of-Flight Mass Spectrometry (MALDI-TOF-MS) have also indicated that conditions of sample handling and preparation are critical [2,3,5,7,8]. Our previous studies and those of others have shown that the combination of LC with Electrospray-Ionization Mass Spectrometry (ESI-MS), abbreviated LC-MS, is suitable to analyze body fluids such as serum or urine [9,10]. The increasing number of applications of LC-MS and LC-MS/MS for the profiling of body fluids or the targeted detection of individual proteins underscores furthermore that this method is capable of achieving concentration sensitivities in the ng-pg/mL range [11-18]. In return, LC-MS provides highly complex data sets when used in the profiling mode (measurement of all detectable compounds in a sample) and it is thus not easy to assess the effect of a given pre-analytical parameter on the overall result.

We describe here an approach to assess the effect of clotting time on LC-MS profiles of serum obtained from a healthy volunteer by univariate and multivariate statistical analysis after data processing. In order to render serum

samples suitable to high-performance LC-MS analysis, proteins were digested with trypsin. Additionally, we investigated the supernatant of acid-precipitated serum samples which are highly enriched in low-molecular weight proteins and peptides (the so-called Peptidome) [3,19-24]. For comparison of samples we used label-free as well as stable-isotope labeling (iTRAQ™) [25] approaches.

2. Materials and Methods

2.1. Description of samples

Serum samples were prepared at the Department of Gynecological Oncology (University Medical Center Groningen, Groningen, The Netherlands) and stored at -80°C in aliquots until analysis. All intermediate fractions that were obtained during sample preparation were stored at -20°C. Glass tubes (Becton Dickinson, #367953), with a separation gel and micronized silica to accelerate clotting, were used for blood collection. Serum was obtained from a single healthy female volunteer, who consented to this study, after different clotting times (1, 2, 4, 6 and 8 h). Serum was prepared by letting the freshly collected blood coagulate at room temperature for 1, 2, 4, 6 or 8 h followed by centrifugation at room temperature for 10 min at 3000 rpm.

2.2. Preparation of serum samples

20 µL of serum were mixed with 80 µL of buffer A (Agilent, Santa Clara, California, USA) of which 80 µL were injected on a Multiple Affinity Removal column (Agilent, 4.6 × 50 mm, Part # 5185-5984) after filtration through a 0.22 µm spin filter (Part # 5185-5990) at 13000 g and 4°C for 10 min to remove particulates. Removal of abundant proteins was performed on a LaChrom HPLC System (Merck Hitachi, www.merck.com) with detection at 280 nm using the following timetable: 0-9 min, 100% buffer A (0.25 mL/min); 9.0-9.1 min, linear gradient 0-100 B % (1 mL/min), 9.1-12.5 min, 100% buffer B (1 mL/min); 12.5-12.6 min, linear gradient 100-0% buffer B (1 mL/min); 12.6-20 min, 100% buffer A (1 mL/min). The flow-through fraction (depleted serum collected between 2-6 min) of a total volume of appr. 1 mL was collected [9]. Each serum sample obtained after different clotting times (1, 2, 4, 6, 8 h) was depleted in duplicate.

Protein concentrations were determined with the Micro BCA™ Protein assay reagent kit (www.piercenet.com) and calculated for an average protein molecular weight of 50 kDa. BSA was used as the calibration standard. Depleted serum samples were digested with trypsin (sequencing grade modified trypsin, Promega, cat # V5111, USA) at an enzyme-to-substrate ratio of 1 : 20 overnight at 37°C with shaking at 400 rpm (Eppendorf Thermomixer) and 4% of the digest were subjected to capillary LC-MS analysis.

2.3. Standard addition of horse heart Cytochrome C

Serum samples were spiked with 21 – 50 pmol of horse heart Cytochrome C (Sigma, www.sigmaaldrich.com, cat. # 9007-43-6) prior to immunodepletion (21 pmol (+) or 50 pmol (++)), of which 10% were subjected to the LC-MS analysis after trypsin digestion. Cytochrome C was alternatively digested with trypsin and added in the same amounts to depleted and trypsin-digested serum (1 h clotting time) prior to LC-MS (21 pmol (n+) or 50 pmol (n++)).

2.4. Cap-LC-MS

All LC-MS analyses were performed on an Agilent 1100 capillary HPLC system coupled on-line to an SL ion-trap mass spectrometer (www.home.agilent.com) equipped with an Atlantis™ dC 18 (1.0 × 150 mm, 3 µm) column that was protected by an Atlantis™ dC 18 in-line trap column (3 µm, 2.1 mm × 20 mm guard column). 40 µL of the pretreated (depleted and digested) fractions corresponding to ~8 µg or 160 pmol of total protein digest (calculated based on a 50 kDa protein) were injected. The autosampler (cat. # G1367A) was equipped with a 100 µL injection loop and a temperature-controlled cooler (cat. # G1330A) maintaining the samples at 4°C. The HPLC system had the following additional components: capillary pump (cat. #, G1376A), solvent degasser (cat. #, G1379A), UV detector (cat. # G1314A) and column holder (cat. #, G1316A). The sample was injected and washed in the back-flush mode for 30 min (0.1% aq. formic acid (FA) and 3% acetonitrile (AcN) at a flow rate of 50 µL/min). Peptides were eluted in a linear gradient from 0 to 70% (0.5%/min) AcN containing 0.1% FA at a flow-rate of 20 µL/min. After each injection, the in-line trap and the analytical column were equilibrated with eluent A (H₂O/AcN/FA; 950:50:1) for 20 min prior to the next injection.

The following settings were used for mass spectrometry during LC-MS. Nebulizer gas: 16.0 psi N₂, drying gas: 6.0 L/min N₂, skimmer: 40.0 V, ionisation voltage: 3500 V, cap. exit: 158.5 V, Oct. 1: 12.0V, Oct. 2: 2.48 V, Oct. RF: 150 Vpp (Voltage, Peak Power Point), Lens 1: -5.0 V, Lens 2: -60.0 V, Trap drive: 53.3, T: 325°C, Scan resolution: enhanced (5500 m/z per second scan speed). Target mass: 600. Scan range: 100-1500 m/z. Spectra were saved in centroid mode. LC-MS chromatographic data were analyzed with Bruker Data Analysis software, version 2.1 (Build 37) [9].

2.5. TCA precipitation of serum samples and MALDI-TOF-MS analysis

TCA, dissolved in 40 µL ice-cold water, was added to 20 µL of the original serum samples to reach a final concentration of 5%. After 30 min on ice, samples were filtered through 0.22 µm spin filters (Part # 5185-5990, Agilent) at 13000 g at 4°C for 10 min to remove particulates. Filtrates were used for further analysis.

For MALDI-TOF-MS analysis, 2 μ L of the filtrates of the TCA precipitate were purified using C18 Stage tips (Proxeon, Odense, DK) according to the manufacturer's instructions. Peptides were eluted in 2 μ L of 5 mg/mL α -Cyano-4-hydroxycinnamic acid (CHCA) in 50% AcN/0.1%TFA and directly spotted on a stainless-steel MALDI target. Analysis was performed in the positive ionization mode using a Voyager DE Pro instrument (Applied Biosystems, Foster City, California, USA). Spectra were acquired in reflectron mode with delayed extraction. Mono-isotopic molecular masses were considered in the further analysis and the instrument was calibrated using singly-charged BSA tryptic fragments with the m/z values of 927.49 Da and 2045.03 Da.

2.6. iTRAQ labeling of depleted serum

Changing of buffer and concentration of samples for iTRAQ labeling of depleted serum were done by ultrafiltration (Concentrators, Spin 5K MWCO, 4 mL, Part no 51855991, Agilent) using 5 mL of 10% AcN with 0.1% TFA in water. Samples were evaporated to dryness in a CentriVapConcentrator (LABCONCO, Kansas City, Missouri, USA) before labeling. iTRAQ labeling was performed according to the manufacturer's protocol [Applied Biosystems, iTRAQ™ Reagents Application Kit-Plasma (Amine-Modifying Labeling Reagents for Plasma Sample Applications)] with modification of the trypsin-to-protein ratio (1:14 instead 1:5.75 w/w). Sequencing grade modified trypsin was from Promega.

2.6.1. Strong Cation Exchange (SCE) fractionation

In order to remove excess iTRAQ reagent and to simplify the ensuing reversed-phase nanoLC-MS-MS analysis, the peptide mixture was washed and fractionated using a strong-cation exchange column [PolyLCinc 4.6 \times 200mm column, column volume: 3.3 mL (Columbia, Maryland, USA)] operated at 0.2 mL/min (AKTA Purifier 10 with frac-900 fraction collector, GE Healthcare Bio-Sciences AB, Uppsala, Sweden). The mobile phase was comprised of two buffers: A: 5 mM $\text{KH}_2\text{PO}_4/\text{H}_3\text{PO}_4$ pH 3, 25% AcN and B: 5 mM $\text{KH}_2\text{PO}_4/\text{H}_3\text{PO}_4$ pH 3, 25% AcN, 1.0 M KCl. The KCl concentration was varied in three segments: 15%B [12 column volumes (CVs)], 50%B (3CVs), 100%B (5CVs), KCl (10 mM/min). The resulting 50 fractions (0.2 mL each) were pooled to obtain 20 fractions based on intensity and dried in a CentriVapConcentrator prior to nanoLC-MS-MS analysis.

2.6.2. Reversed-phase nanoLC-MS-MS

Derivatized peptides (pooled SCE fractions) were dissolved in 50 μ L 2% AcN/0.1% FA and analyzed by nanoLC-MS/MS on a hybrid quadrupole time-of-flight mass spectrometer (QSTAR® XL, Applied Biosystems) connected to an 1100 nanoHPLC system (Agilent). 10 μ L of the SCE fractions were loaded onto a 0.3 mm \times 0.5 cm C18-PepMap trapping column (Dionex, Sunnyvale,

California, USA) at a flow rate of 10 $\mu\text{L}/\text{min}$ (2% AcN/ 0.1% FA). After 30 minutes of washing, the trap column was switched and peptides were separated on a 75 $\mu\text{m} \times 15\text{ cm}$ C18 PepMap column (Dionex). Peptides were eluted at a flow rate of 300 nL/min using a 105 min gradient ranging from 95%A to 50%A [A: H_2O /AcN/FA (950:50:1), B: H_2O /AcN/FA (50/950/1)]. The outlet of the column was connected to an on-line standard coated silica picotip with a 10 μm ID (NewObjective, Woburn, Massachusetts, USA). The typical ionspray voltage was 2200V. Data was acquired using an independent data acquisition (IDA) protocol where, for each cycle, the 3 most abundant multiply-charged peptides (2 to 4 charges) in the MS scan with m/z values between 350 and 1500 amu were selected for MS/MS. For precursor ion selection, a threshold of 30 counts was applied. Each precursor ion was selected twice and then dynamically excluded for the following 60 seconds. Protein identification and quantification was carried out using ProQuant software v1.1 (Applied Biosystems). The search was performed against the Uniprot/SwissProt knowledge database (V49, downloaded May 2005). The search parameters allowed a tolerance up to 0.15 Da for precursor ion selection and the obtained MS/MS fragments, one missed cleavage (trypsin), oxidation of methionine (variable modification) and cysteine modification with iodoacetamide (fixed modification). ProGroup Viewer software v1.0.2 from Applied Biosystem was used to identify proteins with at least 95% confidence. The results obtained from ProGroup Viewer were exported to Microsoft Excel for further analysis.

2.7. iTRAQ labeling of TCA-precipitated serum supernatant

20 μL of TCA-precipitated serum supernatant were evaporated to dryness in a CentriVapConcentrator (LABCONCO) before labeling. iTRAQ labeling was performed according to the manufacturer's protocol with stable-isotopes [iTRAQ reagents with 114 (2 h) and 117 (6 h) isobaric tags] (see also section 2.6 for details). The samples were analyzed by LC-MS-MS with collision parameters optimized for observing the reporter mass tags.

2.8. NanoLC-MS-MS analysis of TCA-precipitated serum

NanoLC-MS/MS was performed as described under 2.6.2 with some modifications. The supernatant of TCA-precipitated serum was diluted ten fold with 2% AcN/ 0.1%FA, and 8 μL of the diluted sample were loaded onto a 0.3 mm \times 0.5 cm C18-PepMap trapping column (Dionex) at a flow rate of 10 $\mu\text{L}/\text{min}$ (2%AcN/0.1% FA). After 3 min of washing, the trap column was switched towards nanoLC and peptides were separated on a 75 $\mu\text{m} \times 15\text{ cm}$ C18 PepMap column (Dionex). Peptides were eluted using a 57 min gradient ranging from 98%A to 50%A (A: H_2O /AcN/FA; 950:50:1, B AcN/ H_2O /FA; 950:50:1). The column was connected to a spray capillary (10 μm tip) (NewObjective). The ion spray voltage applied was 2500 V. Data was acquired

using an independent data acquisition (IDA) protocol where, for each cycle, the most abundant, multiply charged peptides (2 to 4 charges) in the MS scan with m/z between 350 and 1500 were selected for MS/MS. A threshold of 30 counts was applied. Each peptide was selected twice and then dynamically excluded for 15 seconds.

The resulting data was processed using Analyst QS 1.1/BioAnalyst software (Applied Biosystem) with MASCOT (Matrix Science, London, UK) used for protein/peptide identification : MS/MS search parameters: enzyme: none; variable modifications [deamidation (NQ), oxidation(M), Phospho (ST), pyroGlu (N-term E/Q), mono-isotopic, tolerance 0.2 Da]. The search was performed against the Uniprot/SwissProt knowledge database (V49, downloaded May 2005).

3. Data analysis

3.1. Computational environment

For (pre-)processing and multivariate statistical analysis the original Bruker Daltonics label-free LC-MS data files were converted into ASCII-format with the Bruker Data Analysis software. For all other type of data and statistical analyses (e.g. one-way ANOVA), and visualization Matlab (version 7.2.0.232 (R2006a), Mathworks, Natick, Massachusetts, USA) was used. Principal Component Analysis was performed using the PLS toolbox (version 3.5.2, Eigenvector Research Inc., Wenatchee, Washington, USA) under Matlab environment. All data preprocessing work was done on a personal computer equipped with a +3600 MHz AMD processor and with 4 GB of RAM.

3.2. Pre-processing and statistical analysis of label-free LC-MS data

Centroid data were smoothed and reduced using a normalized two-dimensional Gaussian filter, by rounding the nominal m/z ratios to 1 m/z (the original data had a resolution of 0.1 m/z). In the retention time dimension no data reduction was performed. This meshing procedure reduced the number of available data points by roughly a factor 10 and corrected for shifting m/z values as a result of different loadings of the ion-trap during elution of abundant peptides, a phenomenon that is common for ion-trap mass analyzers [26,27]. After meshing the data files, all chromatograms were time-aligned (warped) to a reference data file using Correlation Optimized Warping (COW) [28] based on TICs constructed from signals in the range 100-1500 m/z .

A modified M-N rule was applied for peak detection by first calculating a local median baseline using the sliding window technique separately for each m/z trace. A median window size of 1200 data points, corresponding to 20.84 min, was used with a moving rate of 10 points and the minimum median value was set to 200 counts. According to the M-N rule, a threshold of M-times the local baseline was used and a peak was assigned if, within one m/z trace, the signal exceeded this threshold for at least N consecutive points [26]. For each detected

peak the m/z value, the mean retention times of the three highest measured intensities within the same peak reduced by the local baseline value were stored in a peak list created for every chromatogram. We used a similar approach as Radulovic et al. [26] to obtain optimal settings for M and N. Different values for M (1.5-4) and N (4-8) were applied to two blank LC-MS runs and to two LC-MS runs of depleted, trypsin-digested serum samples. Settings were used at which the ratio between the number of peaks (between 60-155 min) in the samples relative to the blank chromatograms was highest and at which a minimal number of peaks was extracted from the noise in the blank chromatogram (M = 2 and N = 5 in our case).

In order to combine the peak lists from different samples, one-dimensional peak matching was applied using the sliding window technique, in which the same m/z traces were evaluated for peaks that are proximate in time (step size 0.1 min; search window 1.0 min; maximal accepted standard deviation for all retention times within a group of matched peaks was 0.75 min). Missing peak allocation was performed by extracting the background subtracted local signal of the given m/z trace at the given retention time. The generated peak matrix, created from the peak lists of the individual samples, consisted of a peak(row)-sample(column)-intensity(value) matrix. This peak matrix was used for multivariate statistical analysis.

3.3. Analysis of iTRAQ data

The ProQuant software was used to calculate the intensity of 3 reporter ions (m/z : 115, 116 and 117) and to divide them by the intensity of the 4th reporter ion (m/z : 114) for each measured compound. Systematic experimental error occurring during sample preparation and measurement were corrected by subtracting the median of the corresponding ion ratio series (115/114, 116/114 and 117/114) for all compounds and adding one.

The data from two experimental series were analyzed. The first series (Series 1) contained samples with 1 and 2 h of clotting time with each sample measured in duplicate. The second series (Series 2) contained samples with 2, 4, 6 and 8 h of clotting time each of them measured once. Data from Series 1 were normalized to one sample with 1 h clotting time and data in Series 2 were normalized to the 2 h clotting time sample, which was used as denominator in the calculation of ion ratios. The number of identified proteins (95% confidence; see section 2.6.2) was 129 for Series 1 and 96 for Series 2, respectively.

Even after correction for experimental bias between different series of measurements as described above, there is fluctuation of the ion ratios within a single series. The aim of the further statistical analysis is to determine whether a difference in ion ratio is statistically significant relative to the measurement errors (expressed in terms of standard deviations). To this end, all ratios were transformed into decimal logarithms so that all ratios obtained from a single

sample (excluding clotting time-related abundance differences) will have normal distribution. We selected “reference series” to calculate the standard deviation (SD) characterizing a particular set of measurements. In this study the reference series for Series 1 (1 and 2h clotting time measured in duplicate) were the duplicate samples at 1 and 2h, since differences between them cannot be due to clotting time. For Series 2 (2, 4, 6 and 8h clotting time) we based our reference series on the assumption that there are few, if any, differences between 6 and 8h clotting time and that differences between these time points are largely due to experimental error. Ion ratios were plotted against the number of identified proteins (supporting information Fig. S1), Gaussian curves were fitted on the smoothed histograms (histogram between -1 and 1 with 200 steps, smoothing using a Savitzky-Golay algorithm) and the standard deviations (SD) were determined. Proteins with log-transformed ion ratios differing by at least $3 \times \text{SD}$ (99.7% confidence) were considered significantly different from the random fluctuation calculated for the corresponding reference series. resulted in only one compound significantly different from the reference series. For Series 1 the mean of the 2 samples with 2 h of clotting time was used to select proteins that differed significantly different from the reference series. For Series 2 all proteins were included that showed a significant difference for at least one clotting time point. The log-transformed ratios of the selected proteins were used to visualize the clotting time-related differences (see Figure 7).

3.4. Statistical Analysis

For Principal Component Analysis (PCA) the processed data was mean-centered and normalized with respect to the standard deviation. One-way ANOVA was used for univariate statistics.

4. Results and discussion

The preparation of serum is a complex biochemical process and may be difficult to control, since cellular metabolism and the activity of extracellular enzymes continues for many hours after the collection of blood [6,7]. Some authors have noted that clotting time affects the resulting serum proteome, and that these effects are most pronounced in the low-molecular-weight portion, the so-called peptidome [1,3].

It is likely that changes in serum occur due to varying clotting times but it is not clear to which extent such changes affect subsequent proteomics analyses. This is dependent on the employed analytical approach (e.g. shotgun proteomics versus peptidomics) as well as on the use of label-free versus stable-isotope labeling techniques. Finally this depends also on the concentration sensitivity of the employed methodology, since high-abundance proteins may be less prone to significant relative change than those of lower abundance. Since biomarker discovery generally uses a comparative study

design, it will also depend on how large and reproducible group-specific differences in serum proteomes are relative to variation induced by pre-analytical factors such as the clotting time.

In order to study the effect of clotting time on both the serum proteome and the serum peptidome, we have used two complementary methodologies. The first was based on tryptic digestion of serum after depletion of the six most abundant serum proteins followed by LC-MS (proteome), and the second on acid precipitation of serum and analysis of the supernatant (peptidome). The effect of clotting time was assessed with label-free and stable-isotope labeling techniques.

4.1. Label-free serum analysis after depletion and tryptic digestion

Serum was prepared and analyzed by capillary-LC ion-trap mass spectrometry (LC-MS) as previously described [9]. The six most abundant proteins were depleted by immunoaffinity chromatography [29]. In order to compare the effect of clotting time relative to a known change in the serum proteome, horse heart Cytochrome C was added to one group of samples at a concentration of 8 or 20 μ M prior to depletion (the concentration after depletion is 5-times lower due to co-depletion of Cytochrome C). This method was previously shown to have a concentration sensitivity of appr. 0.5 μ M for the added Cytochrome C. Clotting times of 1, 2, 4, 6 and 8 h were compared, in duplicate, by both univariate (ANOVA) and multivariate (PCA) statistical analyses after data processing.

Using in-house-developed algorithms [10], a matched peak matrix was generated from the LC-MS data. This peak matrix contained 12659 aligned features that are defined in terms of retention time, m/z value and intensity. The LC-MS traces (Total Ion Chromatograms; TICs) of depleted, trypsin-digested serum samples showed little if any visible differences (Figure 1).

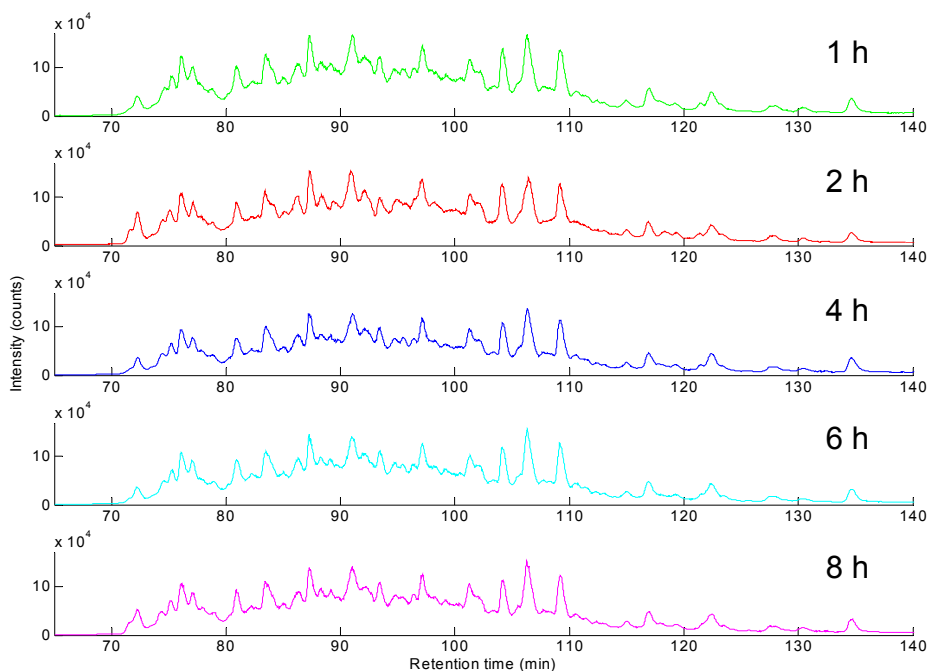


Figure 1. LC-MS analysis of immunodepleted, trypsin-digested human serum. Comparison of Total Ion Chromatograms (TICs) of serum samples obtained after 1, 2, 4, 6 and 8 h of clotting time.

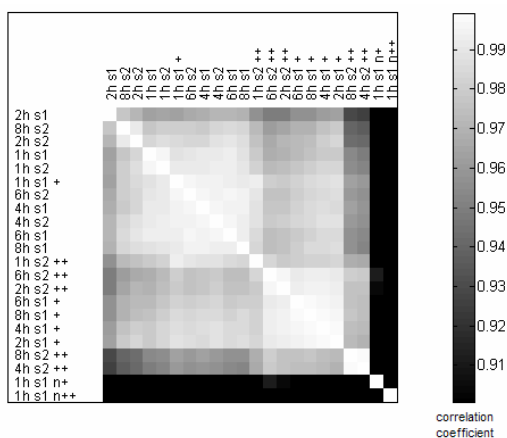


Figure 2. Correlation map of peak lists from all LC-MS analyses of depleted, trypsin-digested serum obtained after different clotting times (1, 2, 4, 6 and 8 h) and spiked with horse heart Cytochrome C (21 pmol (+), 50 pmol (++)) before the depletion step or with a tryptic digest of Horse heart Cytochrome C (21 pmol (n+) and 50 pmol (n++)) before LC-MS analysis. Highest correlation (1.0) = white, lowest correlation (0.9) = black. All non-spiked samples were analyzed in duplicate (s1 and s2) while single measurements were performed on the spiked samples. Samples are grouped by similarity (note that spiking of Cytochrome C results in a clearly decreased correlation coefficient, which is further decreased when a tryptic digest of Cytochrome C is added directly prior to LC-MS analysis).

A correlation map, comparing the peak lists of all LC-MS data files related to samples obtained after different clotting times (highest correlation (1.0) = white, lowest correlation (0.9) = black) confirmed this result by showing an overall high correlation (Figure 2). These results indicate the absence of major differences between serum samples with clotting times ranging from 1 to 8h analyzed by the described method. Contrary to this, the correlation map clearly separated the spiked (before depletion 21pmol (+) or 50pmol (++) and after depletion 21pmol (n+) and 50pmol (n++)) from the non-spiked samples, showing that the effect of spiking of Cytochrome C outweighs the effect of clotting time.

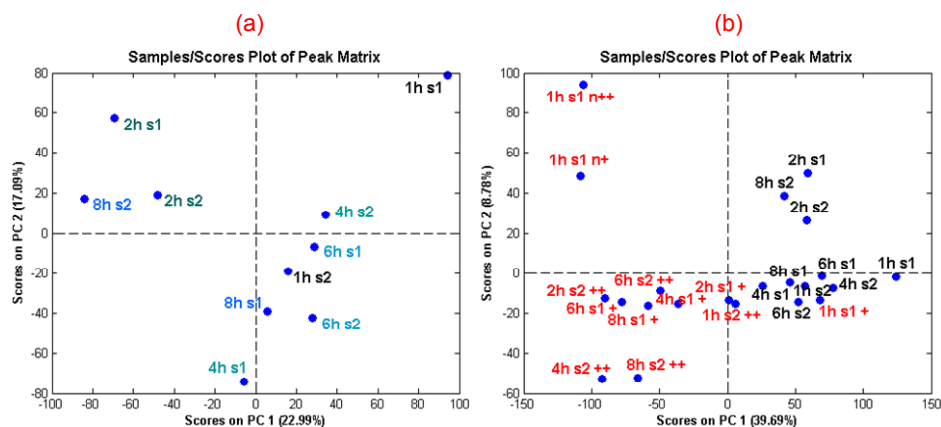


Figure 3. Principal Component Analysis (PCA) score plots of aligned peak matrixes obtained from LC-MS analysis of depleted, trypsin-digested serum after clotting times of 1, 2, 4, 6 and 8h. All analyses were performed in duplicate on different days (denominated as s1 and s2). Samples containing added horse heart Cytochrome C are labeled + (8 μ M) or ++ (20 μ M). In samples labeled n+ (8 μ M) and n++ (20 μ M) a tryptic digest of Cytochrome C was added directly prior to the LC-MS analyses. a) PCA scores plot for non-spiked samples; b) PCA scores plot for spiked (red) and non-spiked (black) samples.

The generated aligned peak matrix was subjected to Principal Component Analysis (PCA) and the calculated scores of PC1 and PC2 were visualized (Figure 3). Fig. 3a shows that there is no clear discrimination between samples with respect to clotting time. Differences between duplicate measurements (s1 and s2) appear to be similar when compared to differences between samples obtained after different clotting times. This indicates that varying the clotting time between 1 and 8h does not lead to detectable differences with the described analytical method. On the contrary, Fig. 3b shows that adding 8 μ M of Cytochrome C (appr. 1.6 μ M after depletion) leads to a clear discrimination between “spiked” (red) and “non-spiked” (black) samples. These data indicate that variable clotting times do not prevent detection of a single protein at a concentration of about 1.5 μ M or 50 μ g/mL (for a 50kDa protein) using the

described methodology. Proteins that are present at this concentration are, however, still considered “classical serum or plasma proteins” [30]. To find smaller local differences between samples that are the result of different clotting times supervised classification methods should be used instead of unsupervised multi-variate statistical methods.

The obtained peak lists were further analyzed by univariate statistical analysis (ANOVA). In this way a p-value is calculated for each peak expressing the magnitude of the effect of a certain factor (clotting time or spiking with Cytochrome C). Table 1 lists several peaks showing significant differences (p-values $< 1.6 \times 10^{-4}$) between the samples obtained at different clotting times and Figure 4 shows box plots for several of the listed peaks. While there is a fairly even distribution of p-values when considering clotting time only, there are more peaks with very low p-values in the case of added Cytochrome C (p-values $< 2.2 \times 10^{-11}$; Table 2) indicating that addition of Cytochrome C clearly outweighs the effect of clotting time.

The peaks shown in Table 1 were subjected to LC-MS-MS analysis and identified as corresponding to fragments of Fibrinogen alpha chain (FIBA_HUMAN (P02671); m/z = 733 and 809) or to Fibrinogen beta chain (FIBB_HUMAN (P02675); m/z = 1325) with a pyroglutamic acid at the N-terminus (see Fig. S2-S4, supporting information for further details), while all peaks in Table 2 were derived from Cytochrome C. Extracted Ion Chromatograms (EICs) of discriminative peaks corresponding to fragments of Fibrinogen alpha chain (m/z: 733, 91.22 min and m/z: 809, 93.42 min) at 1, 2, 4, 6 and 8 h of clotting times are presented in Figure S5 (supporting information).

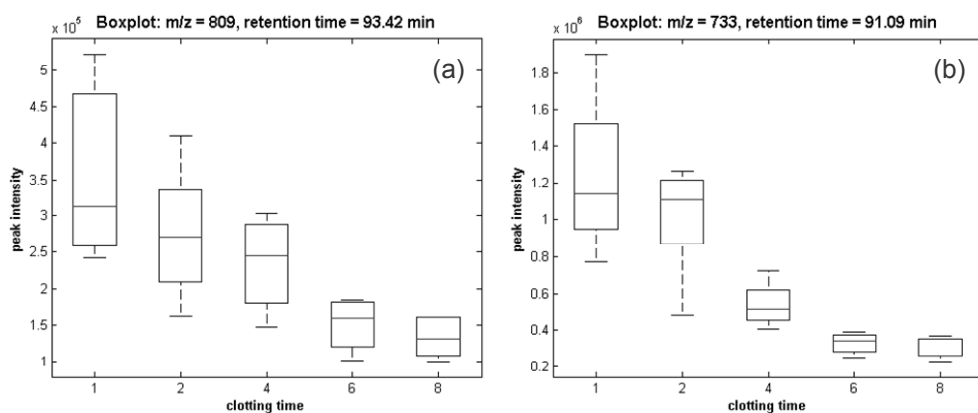


Figure 4. Box plots of the intensity values of two fibrinogen-derived peptide peaks that change significantly with clotting time (see Table 1). Peptides were identified as fragments of Fibrinogen alpha chain [(a) DSGEGDFLAEGGGVR; (b) ADpSGEGDFLAEGGGVR); FIBA_HUMAN (P02671)] (see Fig. S2 and S3, supporting information).

The peaks which correspond to p-values below 1.6×10^{-4} with respect to clotting time are listed in Table 1 and those with respect to Cytochrome C addition (p-values below 2.2×10^{-11}) in Table 2.

Table 1. Peaks detected in serum that are significantly affected by clotting times ranging from 1 to 8h. Calculations are based on a comparison of intensities by one-way ANOVA.

p-value	m/z	Retention time (min)	Peptide sequence
$2.6 \cdot 10^{-7}$	733 ¹	91.09	DSGEGDFLAEGGGVR
$1.1 \cdot 10^{-4}$	1325 ²	103.87	QGVNDNEEGFFS
$1.6 \cdot 10^{-4}$	809	93.42	ADSGEGDFLAEGGGVR

¹ isotopic peaks at m/z 732 ($p = 1.4 \times 10^{-4}$) and 734 ($p = 6.2 \times 10^{-8}$) show also significant differences.

² the isotopic peak at m/z 1324 shows also a significant difference ($p = 2.8 \times 10^{-4}$).

Table 2. Peaks detected in serum that are significantly affected by the addition of horse heart Cytochrome C ($8\mu\text{M}$ prior to depletion) independent of clotting times ranging from 1 to 8h. Calculations are based on a comparison of intensities by one-way ANOVA.

m/z measured ¹	m/z expected	charge state	retention time (min)	p-value	Cytochrome C sequence
634	634,4	+1	76.14	$<1.00\text{e-}015$	IFVQK
453	454,3	+2	92.18	$1.74\text{e-}014$	MIFAGIKK
966	964,5	+1	98.67	$9.16\text{e-}014$	EDLIAYLK
559	559,3	+3	99.22	$9.51\text{e-}014$	GGKHKTGPNLHGLFGR
452	450,9	+3	92.22	$2.30\text{e-}013$	TEREDLIAYLK
749	748,3	+2	104.36	$4.57\text{e-}013$	EETLMEYLENPK
782	781,4	+2	95.24	$4.61\text{e-}013$	HKTGPNLHGLFGRK
798	799,9	+2	103.19	$7.82\text{e-}012$	KTGQAPGFTYTDANK
1169	1168,6	+1	91.82	$1.60\text{e-}011$	TGPNLHGLFGR

¹ after meshing to integer values

4.2. Detailed analysis of Fibrinopeptides after TCA precipitation

The analysis of depleted and trypsin-digested serum indicated that fibrinogen-related peptides were affected by clotting time, as might be expected. We therefore examined the effect of clotting time on the overall serum peptidome in more detail. TCA precipitation after 2 and 6 hours clotting time was used to enrich the low molecular weight fraction and the resulting supernatant was analyzed by MALDI-TOF-MS to gain an overview of the peptide patterns (Figure 5).

A small set of prominent peptide peaks was observed. The spectrum was dominated by a peak at m/z 1465.8, which was identified as an N-terminally truncated form of Fibrinopeptide A (FPA) (DSGEGDFLAEGGGVR, doubly-charged ion of m/z 733.8 (supporting information, Fig. S2, Table S1)) by LC-MS/MS. This peptide was also shown to be most significantly affected by clotting time based on the univariate statistical analysis of the LC-MS data (see Table 1). A minor signal for intact FPA (ADSGEGDFLAEGGGVR) at m/z 1536.6 was also observed. Database-search of the LC-MS/MS data showed that the majority of the observed signals using this approach were derived from FPA (Table S1). Interestingly, two post-translational modifications of FPA were seen in the MALDI spectrum. Peaks at m/z 1545.7 and 1616.8 correspond to the phosphorylated forms DpSGEGDFLAEGGGVR (doubly-charged ion m/z 773.4) and ADpSGEGDFLAEGGGVR (doubly-charged ion m/z 808.8) as confirmed by LC-MS/MS (supporting information, Fig. S3,S4). The MALDI spectrum showed also two less well-resolved peaks at m/z 1524.5 and m/z 1453 (Figure 5C). These peaks represented metastable ions due to the formation of dehydro-alanine (loss of phosphoric acid from the phosphorylated serine during flight to the detector). Besides FPA-related peptides, partial sequences of Fibrinopeptide B (FPB) were also identified.

In order to gain further insight into the ratios of FPA- and FPB-related peptides at 2 and 6h clotting time, supernatants after TCA precipitation were labeled with stable-isotopes (iTRAQ reagents with 114 (2h) and 117 (6h) isobaric tags). The samples were analyzed by LC-MS-MS with collision parameters optimized for observing the reporter mass tags. Database search of the MS/MS data sets allowed the identification of a greater variety of peptides than in the original, targeted analyses. Fragments of FPA and FPB were readily identified but quantification was problematic, due to the fact that the software (proQuantTM 1.1) is designed for tryptic peptides and not for naturally occurring peptides with non-tryptic cleavage sites. Extracted ion traces related to Fibrinopeptides from the data base search were therefore inspected manually. As an examples, the MS/MS spectra of the iTRAQ-labeled peptides ADpSGEGDFLAEGGGVR and FLAEGGGV are shown in Figure 6. In the insert the iTRAQ reporter ions indicate that the amount of this peptide is reduced by a factor 1.75 between 2 and 6h clotting time (Figure 6A), whereas another iTRAQ-labeled FPA peptide (FLAEGGGV; m/z 447 doubly-charged ion) showed a 2.8-fold increase over this time period (Figure 6B). An overview of all identified peptides is given in Table S2 (supporting information).

These experiments show that a number of peptides derived from Fibrinogen changed in abundance with respect to clotting time and that these peptides can be easily detected in the so-called peptidome due to their relatively high-abundance.

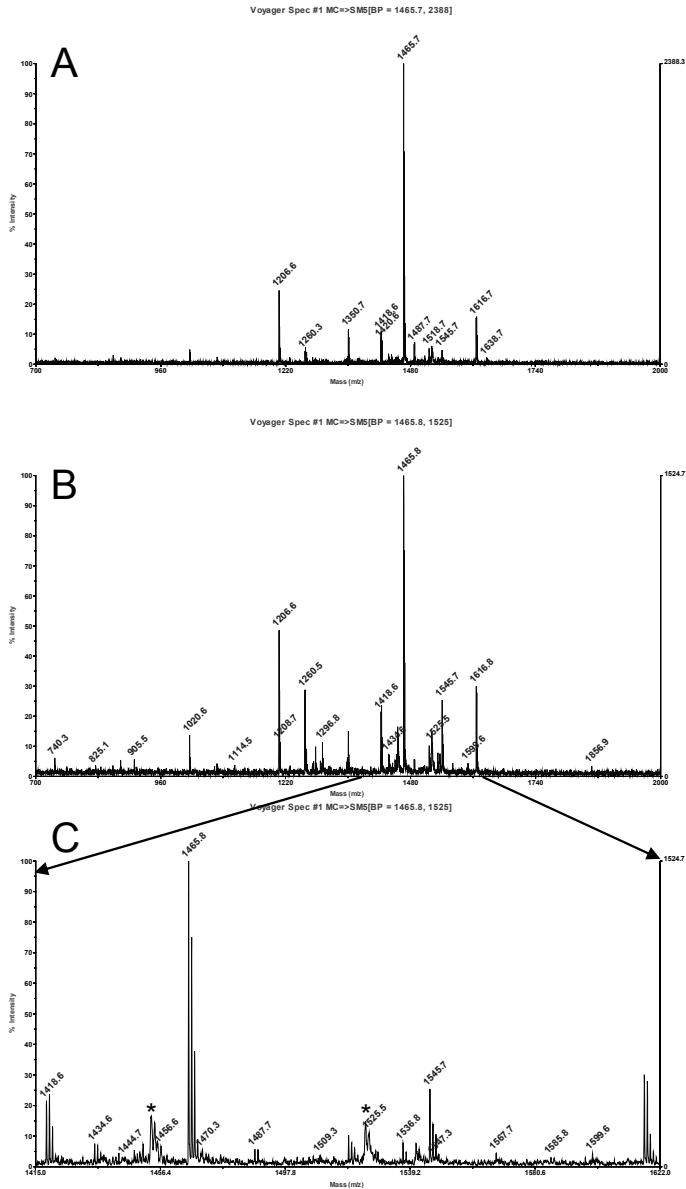


Figure 5. MALDI-TOF-MS spectrum of the supernatant of TCA-precipitated serum after 2 (A) and 6h (B) clotting time. The major peak at 1465.8 corresponds to an N-terminally truncated form of Fibrinopeptide A (FPA) and the minor peak at 1536.8 (Fig. 5C) corresponds to full-length FPA. These forms of FPA were also observed as serine-phosphorylated forms (m/z 1545.7 and m/z 1616.8). (C) Metastable ions at m/z 1453 and 1524.5 (labeled with * in panel (C)) indicate the presence of dehydro-alanine due to decay of the metastable phosphoserine residue during MALDI-TOF-MS analysis.

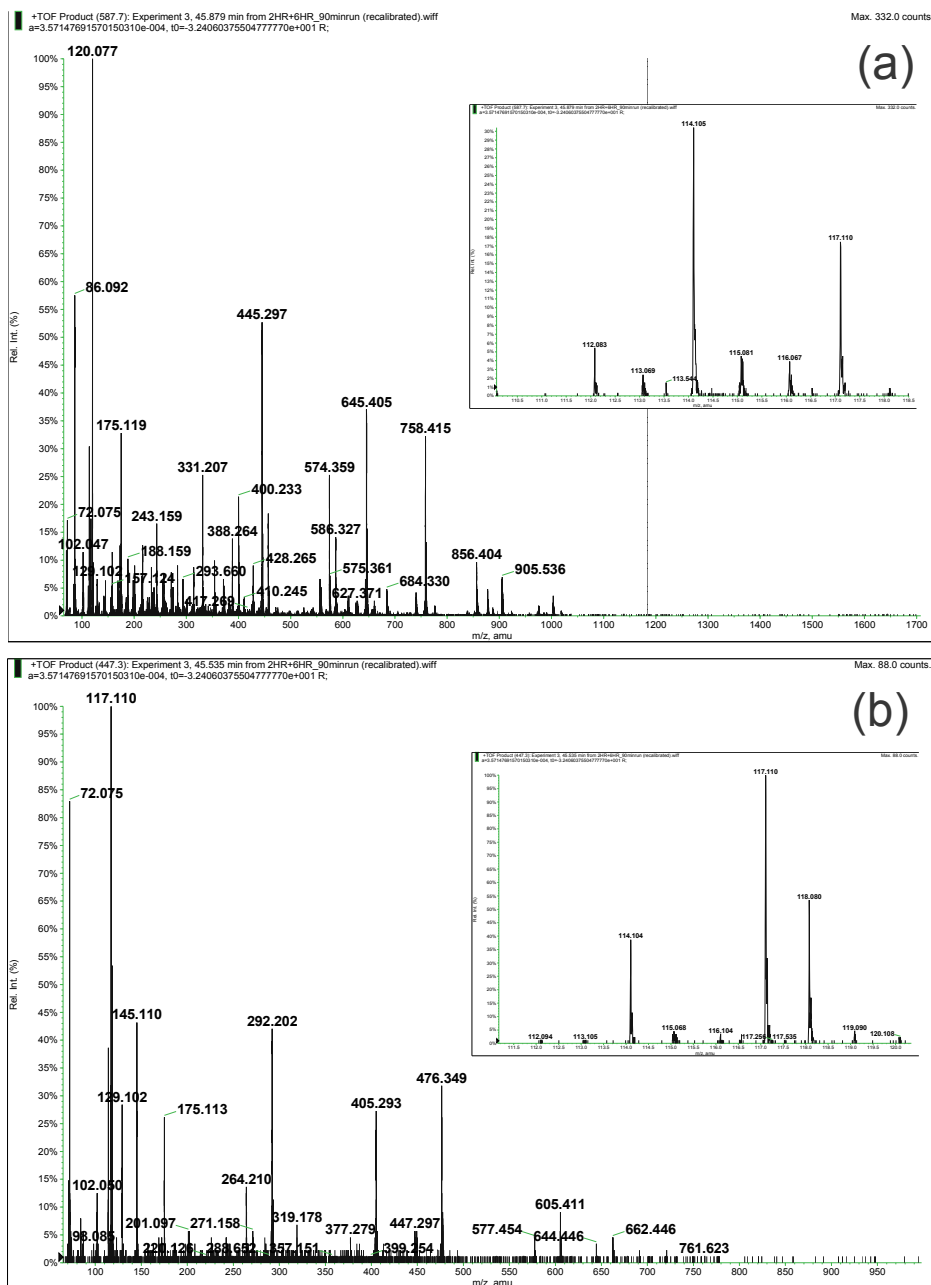


Figure 6. MS/MS spectrum of the iTRAQ-labeled, phosphorylated Fibrinopeptide A (FPA) (ADpSGEGDFLAEGGGVR) (a). The insert shows the relative abundance of this peptide at 2h [114 mass tag] and 6h [117 mass tag] clotting time (ratio 2h/6h = 1.75). MS/MS spectrum of iTRAQ-labeled truncated FPA (FLAEGGGV; m/z 447.3 doubly-charged ion) which increases by 2.8-fold between 2 [mass tag 114] and 6h [mass tag 117] clotting time (b).

4.3. Multiplexed analysis of trypsin-digested serum

Since the concentration sensitivity of the label-free analytical method described in section 4.1 reaches only about 0.5 μ M, it was of interest to apply a two-dimensional HPLC method with stable-isotope labeling to obtain a more comprehensive view of the effect of clotting time on the serum proteome. The initial experimental design consisted of varying clotting time between 2, 4, 6 and 8 h followed by immunodepletion of the six most abundant proteins. Proteins were furthermore reduced and alkylated and digested with trypsin prior to derivatization with a different isobaric tag for each clotting time [114 (2h), 115 (4h), 116 (6h) and 117 (8h)]. The combined samples (50 μ g each) were fractionated by Strong Cation-Exchange HPLC (SCX; 20 fractions) prior to reversed-phase nanoLC-MS-MS analysis. The LC-MS-MS data were used for relative quantification and identification of proteins. 96 proteins were identified with a confidence higher than 95% and quantified relative to each other using the reporter mass tags in the MS/MS spectra of selected multiply-charged ions. Examples of spectra produced after iTRAQ labeling are shown (supporting information, Figure S6).

All the data files were analyzed separately using the ProQuant software. This software package is able to quantify the reporter ions and to calculate the ratio relatively to a preset denominator, in this case the 2h clotting time point (mass tag 114). A bias (multiplication factor) was applied in order to correct for systematic differences in the iTRAQ ratios due to experimental errors related to sample preparation (e.g. combining the different reaction mixtures; small differences in dilutions). The software calculates this bias and corrects the ratios assuming that the majority of proteins does not differ in response to clotting time and that the median iTRAQ ratio should be 1. In the experiment used the mean bias applied was 1 for the 114 (2h) signal (was set to 1), 1.09 for 115 (4h), 1.34 for 116 (6h) and 1.14 (8h) for 117.

Visualization of the obtained ratios for the 96 identified proteins (Figure 7a) showed that 10 proteins have significant differences in ratios relative to the 2h clotting time point (Figure 7a). Three of them (Complement component C8 alpha chain precursor (P07357), Hepatocyte growth factor activator precursor (HGF activator) (Q04756), Ig kappa chain C region (P01834)) were increased and seven (Insulin-like growth factor II precursor (IGF-II) (Somatomedin A) (P01344), Leucine-rich alpha-2-glycoprotein precursor (LRG) (P02750), Complement factor H-related protein 1 precursor (FHR-1) (H factor-like protein 1) (H-factor-like 1) (Q03591), Serum amyloid P-component precursor (SAP) (9.5S alpha-1-glycoprotein) ((P02743), Coagulation factor IX (P00740), Complement factor H-related protein 2 precursor (P36980) and Apolipoprotein C-IV precursor (P55056)) were decreased. All mentioned proteins were increased or decreased relative to the reference series (see Materials and Methods; section 3.3) with a confidence above 99.7%. However, none of the mentioned proteins was more than +/- 50% increased or decreased.

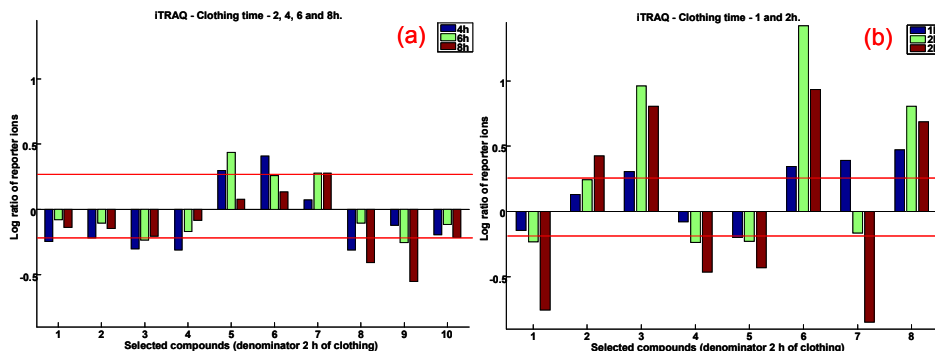


Figure 7. Compound showing significant difference in sample having clotting time of 2, 4, 6 and 8 hours (a) and clotting time 1 and 2 hours (in duplicate) (b) using iTRAQ quantification. The number on the figures are corresponding to the following proteins: (a): 1: SAMP_HUMAN (P02743) 2: FA9_HUMAN (P00740), 3: FHR2_HUMAN (P36980), 4: APOC4_HUMAN (P55056), 5: CO8A_HUMAN (P07357), 6: HGFA_HUMAN (Q04756), 7: KAC_HUMAN (P01834), 8: IGF2_HUMAN (P01344), 9: A2GL_HUMAN (P02750), 10: FHR1_HUMAN (Q03591); (b): 1: APOC1_HUMAN (P02654), 2: FIBA_HUMAN (P02671), 3: SC1_HUMAN (Q08554), 4: APOF_HUMAN (Q13790), 5: A2GL_HUMAN (P02750), 6: ALBU_HUMAN (P02768), 7: EPC2_HUMAN (Q52LR7), 8: HV3P_HUMAN (P01777).

Since it was not possible, at the time, to include more than 4 different clotting time points in one iTRAQ experiment, we compared the effects of shorter clotting times [1h (mass tags 114, 115) and 2 h (mass tags 116, 117)] in separate duplicate experiments. The analysis resulted in the identification of 129 proteins at the 95% confidence level. Based on an analogous analysis of the obtained data, eight proteins were shown to differ between serum samples with 1h as compared to 2h clotting time (Fig. 7b). Human serum albumin precursor (P02768), Fibrinogen alpha chain (P02671), Desmocollin-1 (Q08554) and Ig heavy chain V-III region TEI (P01777) were increased in concentration (up to +70%) and Apolipoprotein C-I (P02654), Apolipoprotein F (Q13790), Leucine-rich alpha-2-glycoprotein (P02750) and Enhancer of polycomb homolog 2 (Q52LR7) were decreased by at least 100%. Overall it appears that changes to the serum proteome are more drastic between 1 and 2 h clotting time and that the serum proteome “stabilizes” after 2 h.

Fibrinogen alpha chain (P02671), from which Fibrinopeptide A is cleaved off during blood clotting, was up-regulated after 2h clotting time compared to 1h but this difference was not significant at a confidence level of 99.7%. Fibrinogen alpha chain is involved in blood clotting and Fibrinopeptide A is the product of Fibrinogen cleavage during coagulation. Since we did not observed further changes in the relative abundance of this peptide after 2h clotting time, it is likely that coagulation was completed (Figure 7b).

Coagulation factor IX (P00740), consisting of Coagulation factor IXa light chain and Coagulation factor IXa heavy chain, was down-regulated after two

hours clotting compare to 4, 6 and 8 hours but only in -0.1/-0.2 times which fall into the the normal measurement fluctuation. Leucine-rich alpha-2-glycoprotein (P02750) was down-regulated in all samples compared to one hour clotting sample. Leucine-rich alpha-2-glycoprotein is known as marker for granulocytic differentiation [31] which is involved in formation of platelets. Platelets are components of blood coagulation and activate the final step in the process of blood coagulation, thus Leucine-rich alpha-2-glycoprotein related to blood coagulation cascade.

Our studies with immunodepleted, trypsin-digested serum indicate that the serum proteome is rather stable once clotting has proceeded for at least 2h. Univariate statistical comparisons of the processed LC-MS data showed that fibrinogen-derived peptides change significantly with clotting time (p-values 1.6×10^{-4} to 7×10^{-8}) but that most other peptides remain stable in their intensities (+/- 25%). A more detailed analysis of these peptides revealed that they form a family of truncated forms, some of which contain post-translational modifications such as phospho-serine or pyroglutamic acid. Stable-isotope labeling combined with a two-dimensional chromatographic separation showed that 10 out of 96 identified proteins showed statistically significant differences between 2, 4, 6 and 8h clotting time. Our results indicate that serum stabilizes after 2h clotting under the described conditions, while shorter clotting times need to be more tightly controlled. This opens the possibility of using disease-relevant serum samples for biomarker discovery that would otherwise be very difficult to generate de novo.

Other authors have observed that clotting time needs to be tightly controlled when focusing on the so-called peptidome, the low-molecular weight complement of the proteome [3,5]. This is confirmed by our results, which show that especially fibrinogen-related peptides vary greatly with clotting time. These fibrinopeptides form a family of related molecules that are N- or C-terminally truncated as well as phosphorylated or containing N-terminal pyro-glutamic acid residues (see Table S2, supporting information). An increasing number of authors report that proteolytic degradation products of high-abundance proteins may be related to cancer and may thus be of interest as biomarkers [32,33]. This interesting observation was based on a rigorously controlled sample handling and preparation protocol, which may be difficult to implement in routine clinical diagnostic laboratories [8]. Some of these biomarker candidates were forms of Fibrinopeptides, which are also affected by clotting time. It will be interesting to follow the development of this field of biomarker research, since cancer development is often associated with changes in the balance between the proteolytic systems of the coagulation cascade, the fibrinolytic and the complement system. Villanueva et al. [33] assume that these major proteolytic systems provide the “founder peptides” that are subsequently further degraded by cancer-specific proteases.

Our study sheds additional light on the effect of clotting time on the composition of the serum proteome and peptidome. Biomarker discovery is relying on changes in protein or peptide abundance related to disease relative to unavoidable changes due to biological and analytical variations. Abundance ratios of some 100 identified proteins changed no more than 10 between 2 and 8h clotting time while changes were 8 between 1 and 2h. Most analyzed proteins were not significantly affected by clotting time. While the expected changes in putative, disease-relevant biomarkers can not be predicted, it is possible to make a calculation as to how large the relative difference should be in order to be statistically significant. The kinetics of blood coagulation are affected by multiple factors related to the blood collection tubes (e.g. some tubes contain clotting activators) as well as to the medication a patient might be taking (e.g. use of anti-coagulants, chemotherapy). We did not study these parameters but it is clear that they must be taken into account in comparative biomarker discovery studies. The detected fibrinopeptides might be good indicators to assess the clotting conditions a posteriori and thus to define inclusion/exclusion criteria for serum samples in existing biobanks. The question whether serum or plasma should be the preferred blood derivative for a given study cannot be answered definitively. Both types of samples have their advantages and disadvantages and plasma has also been shown to be dependent on a number of parameters, notably the anti-coagulant [3].

References

- [1]. Drake RR, Cazares LH, Corica A, Malik G, Schwegler EE, Libby AE, Wright GL, Semmes OJ, Adam BL. Quality control, preparation and protein stability issues for blood serum and plasma used in biomarker discovery and proteomic profiling assays. *Bioprocessing Journal* 2004;43-49.
- [2]. Hsieh SY, Chen RK, Pan YH, Lee HL. Systematical Evaluation of the Effects of Sample Collection Procedures on Low-Molecular-Weight Serum/Plasma Proteome Profiling. *Proteomics* 2006;6(10):3189-3198.
- [3]. Tammen H, Schulte I, Hess R, Menzel C, Kellmann M, Mohring T, Schulz-Knappe P. Peptidomic Analysis of Human Blood Specimens: Comparison Between Plasma Specimens and Serum by Differential Peptide Display. *Proteomics* 2005;5:3414-3422.
- [4]. Timms JF, Arslan-Low E, Gentry-Maharaj A, Luo Z, T'jampens D, Podust VN, Ford J, Fung ET, Gammerman A, Jacobs I, Menon U. Preanalytic Influence of Sample Handling on SELDI-TOF Serum Protein Profiles. *Clin. Chem.* 2007;53(4):645-656.
- [5]. Villanueva J, Philip J, Chaparro CA, Li Y, Toledo-Crow R, DeNoyer L, Fleisher M, Robbins RJ, Tempst P. Correcting Common Errors in Identifying Cancer-Specific Serum Peptide Signatures. *J. Proteome. Res.* 2005;4:1060-1072.
- [6]. Ekblad L, Baldetorp B, Ferno M, Olsson H, Bratt C. In-Source Decay Causes Artifacts in SELDI-TOF MS Spectra. *J. Proteome Res.* 2007;6(4):1609-1614.
- [7]. Aristoteli LP, Molloy MP, Baker MS. Evaluation of Endogenous Plasma Peptide Extraction Methods for Mass Spectrometric Biomarker Discovery. *J. Proteome Res.* 2007;6:571-581.
- [8]. Villanueva J, Philip J, Entenberg D, Chaparro CA, Tanwar MK, Holland EC, Tempst P. Serum Peptide Profiling by Magnetic Particle-Assisted, Automated Sample Processing and MALDI-TOF Mass Spectrometry. *Anal. Chem.* 2004;76:1560-1570.
- [9]. Govorukhina NI, Reijmers TH, Nyangoma SO, van der Zee AGJ, Jansen RC, Bischoff R. Analysis of Human Serum by Liquid Chromatography-Mass Spectrometry: Improved Sample Preparation and Data Analysis. *Journal of Chromatography A* 2006;1120:142-150.
- [10]. Kemperman RF, Horvatovich PL, Hoekman B, Reijmers TH, Muskiet FA, Bischoff R. Comparative Urine Analysis by Liquid Chromatography-Mass Spectrometry and Multivariate Statistics: Method Development, Evaluation, and Application to Proteinuria. *J. Proteome Res.* 2007;6:194-206.
- [11]. Aguiar M, Masse R, Gibbs BF. Mass Spectrometric Quantitation of C-Reactive Protein Using Labeled Tryptic Peptides. *Analytical Biochemistry* 2006;354:175-181.
- [12]. Berna MJ, Zhen Y, Watson DE, Hale JE, Ackermann BL. Strategic Use of Immunoprecipitation and LC/MS/MS for Trace-Level Protein Quantification: Myosin Light Chain 1, a Biomarker of Cardiac Necrosis. *Anal. Chem.* 2007 Jun 1;79(11):4199-205.
- [13]. Bondar OP, Barnidge DR, Klee EW, Davis BJ, Klee GG. LC-MS/MS Quantification of Zn-{Alpha}2 Glycoprotein: A Potential Serum Biomarker for Prostate Cancer. *Clinical Chemistry*;2007;53(4):673-678.
- [14]. Lin S, Shaler TA, Becker CH. Quantification of Intermediate-Abundance Proteins in Serum by Multiple Reaction Monitoring Mass Spectrometry in a Single-Quadrupole Ion-trap. *Anal. Chem.* 2006;78:5762-5767.
- [15]. Liu T, Qian WJ, Gritsenko MA, Xiao W, Moldawer LL, Kaushal A, Monroe ME, Vamum SM, Moore R. J.; Purvine, S. O.; Maier, R. V.; Davis, R. W.; Tompkins, R. G.; Camp, D. G.; Smith RD. High Dynamic Range Characterization of the Trauma Patient Plasma Proteome. *Mol. Cell Proteomics.* 2006;5(11):2167-2174.
- [16]. Qian WJ, Jacobs JM, Liu T, Camp DG, Smith RD. Advances and Challenges in Liquid Chromatography-Mass Spectrometry Based Proteomic Profiling for Clinical Applications. *Mol. Cell Proteomics.* 2006;5(10):1727-1744.
- [17]. Shen Z, Want EJ, Chen W, Keating W, Nussbaumer W, Moore R, Gentle TM, Siuzdak G. Sepsis Plasma Protein Profiling With Immunodepletion, Three-Dimensional Liquid Chromatography Tandem Mass Spectrometry, and Spectrum Counting. *J. Proteome. Res.* 2006;5:3154-3160.

- [18]. Wang G, Wu WW, Zeng W, Chou CL, Shen RF. Label-Free Protein Quantification Using LC-Coupled Ion-trap or FT Mass Spectrometry: Reproducibility, Linearity, and Application With Complex Proteomes. *J. Proteome. Res.* 2006;5:1214-1223.
- [19]. Petricoin EF, Belluco C, Araujo RP, Liotta LA. The Blood Peptidome: a Higher Dimension of Information Content for Cancer Biomarker Discovery. *Nat. Rev. Cancer.* 2006;12:961-967.
- [20]. Raida M, Schulz-Knappe P, Heine G, Forssmann WG. Liquid Chromatography and Electrospray Mass Spectrometric Mapping of Peptides From Human Plasma Filtrate. *Journal of the American Society for Mass Spectrometry* 1999;10:45-54.
- [21]. Schrader M, Schulz-Knappe P. Peptidomics Technologies for Human Body Fluids. *Trends in Biotechnology* 2001;19:555-60.
- [22]. Schulz-Knappe P, Schrader M. Peptidomics in Biomarker and Drug Discovery. *Current Drug Discovery* 2003;21-24.
- [23]. Schulz Knappe P, Raida M, Meyer M, Quellhorst EA, Forssmann WG. Systematic Isolation of Circulating Human Peptides: the Concept of Peptide Trapping. *Eur J Med Res* 1996;1:223-36.
- [24]. Schulz Knappe P, Schrader M, Standker L, Richter R, Hess R, Jurgens M, Forssmann WG. Peptide Bank Generated by Large-Scale Preparation of Circulating Human Peptides. *J Chromatogr A* 1997;776:125-132.
- [25]. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlett-Jones M, He F, Jacobson A, Pappin DJ. Multiplexed Protein Quantitation in *Saccharomyces Cerevisiae* Using Amine-Reactive Isobaric Tagging Reagents. *Mol Cell Proteomics* 2004;3:1154-1169.
- [26]. Radulovic D, Jelveh S, Ryu S, Hamilton TG, Foss E, Mao Y, Emili A. Informatics Platform for Global Proteomic Profiling and Biomarker Discovery Using Liquid Chromatography-Tandem Mass Spectrometry. *Mol. Cell Proteomics.* 2004;3:984-997.
- [27]. Hao C, March RE. A Survey of Recent Research Activity in Quadrupole Ion-trap Mass Spectrometry. *International Journal of Mass Spectrometry* 2001;212:337-357.
- [28]. Tomasi GF. v. d. BCA. Correlation Optimized Warping and Dynamic Time Warping As Preprocessing Methods for Chromatographic Data. *Journal of Chemometrics* 2004;18:231-241.
- [29]. Dekker LJ, Bosman J, Burgers PC, van Rijswijk A, Freije R, Luidert T, Bischoff R. Depletion of High-Abundance Proteins From Serum by Immunoaffinity Chromatography: A MALDI-FT-MS Study. *Journal of Chromatography B* 2007;847:65-69.
- [30]. Anderson NL, Anderson NG. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Mol Cell Proteomics* 2002;1:845-867.
- [31]. O'Donnell LC, Druhan LJ, Avalos BR. Molecular characterization and expression analysis of leucinerich α -2-glycoprotein, a novel marker of granulocytic differentiation. *Journal of Leukocyte Biology* 2002;72:478-485.
- [32]. Overall CM, Dean RA. Degradomics: Systems Biology of the Protease Web. Pleiotropic Roles of MMPs in Cancer. *Cancer Metastasis Rev.* 2000;25:69-75.
- [33]. Villanueva J, Shaffer DR, Philip J, Chaparro CA, Erdjument-Bromage H, Olshen AB, Fleisher M, Lilja H, Brogi E, Boyd J, Sanchez-Carbayo M, Holland EC, Cordon-Cardo C, Scher HI, Tempst P. Differential Exoprotease Activities Confer Tumor-Specific Serum Peptidome Patterns. *J. Clin. Invest.* 2006;116:271-284.

Supplementary material

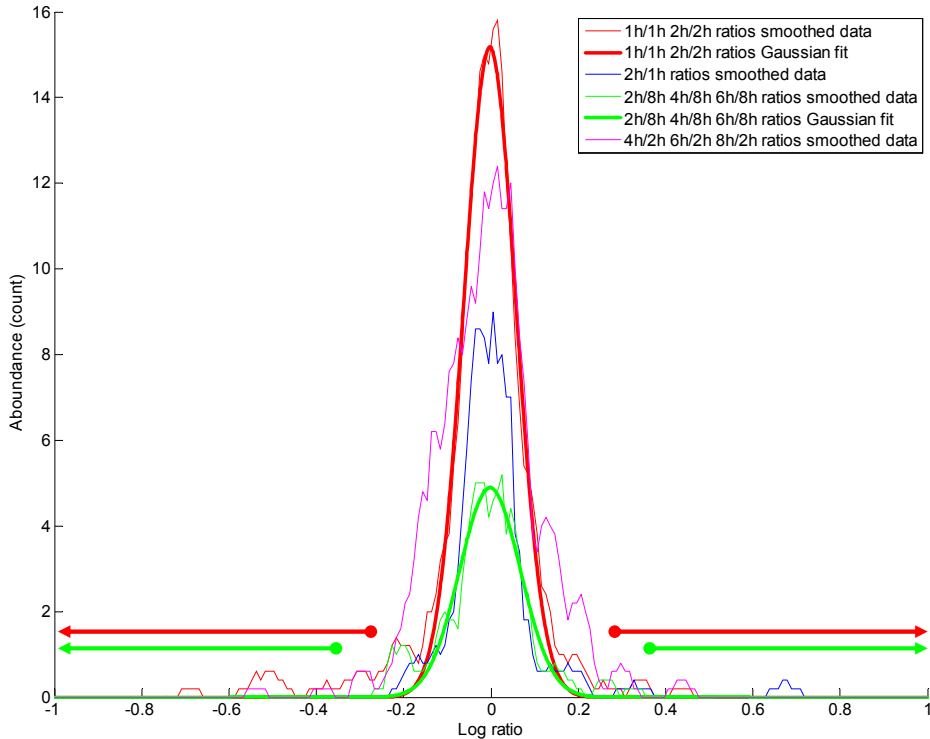


Figure S1. Evaluation of iTRAQ data. The thin red and green are the smoothed logtransformed ratio of the references sets after correction with the median, and the bold corresponding lines are the Gaussian fit. The thin blue and purple curve is the ratios corresponding to set 1 and 2 in which significance peak ratios has to be determined. The arrows are indicating the log ratio of area considered as significant 3 SD (99.7%).

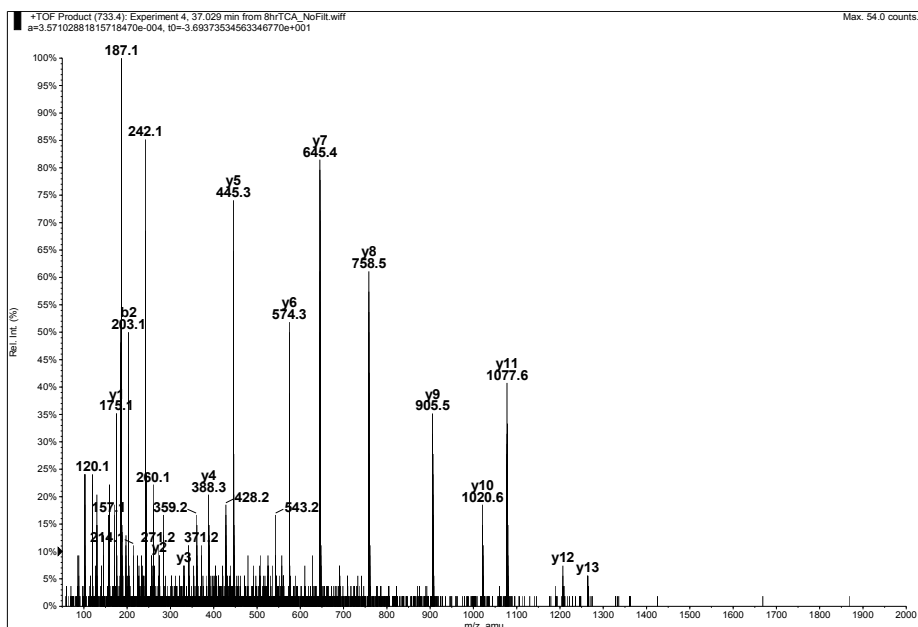


Figure S2. MS-MS spectrum of a FPA fragment (DSGEGDFAEGGGVR; m/z 733.4, doubly charged)

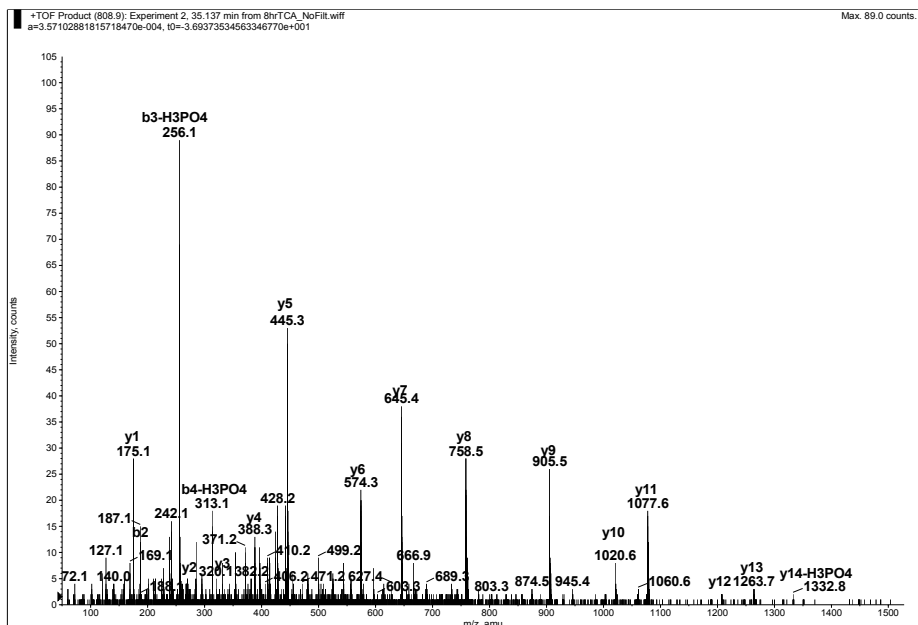


Figure S3. MS/MS spectrum of a phosphorylated FPA fragment (ADpSGEGDFLAEGGGVR; m/z 808.9, doubly-charged).

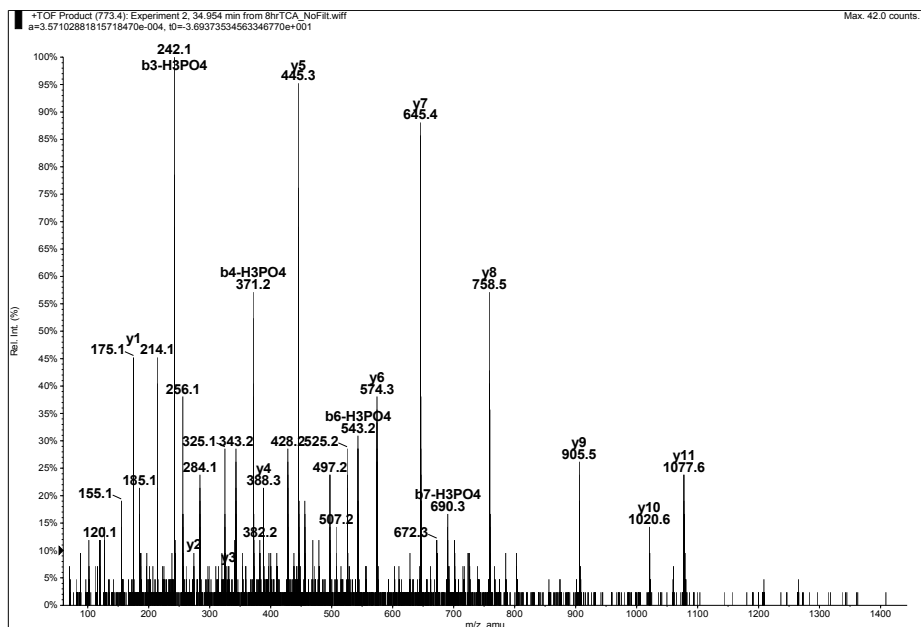


Figure S4. MS-MS spectrum of a phosphorylated FPA fragment m/z 773.4 (MM1464.8, DpSGEGDFLAEGGGVR, doubly charged)

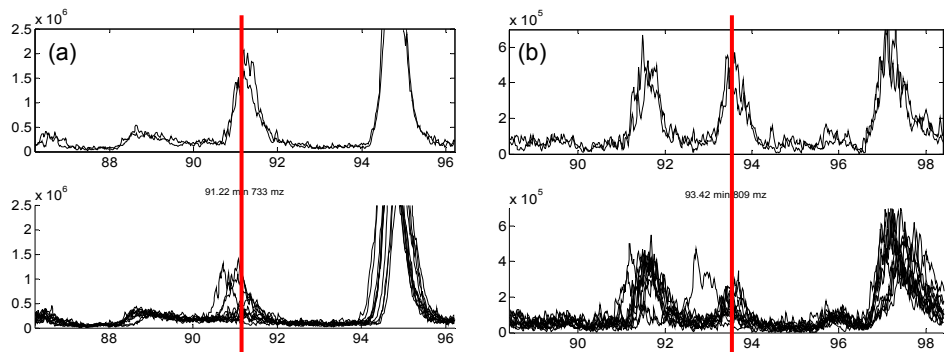


Figure S5. Discriminative peaks [733 m/z , 91.22 min (a) and 809 m/z and 93.42 min (b)] between 1 (upper traces) and 2 (lower traces), 4, 6 and 8 h (lower traces) of clotting time.

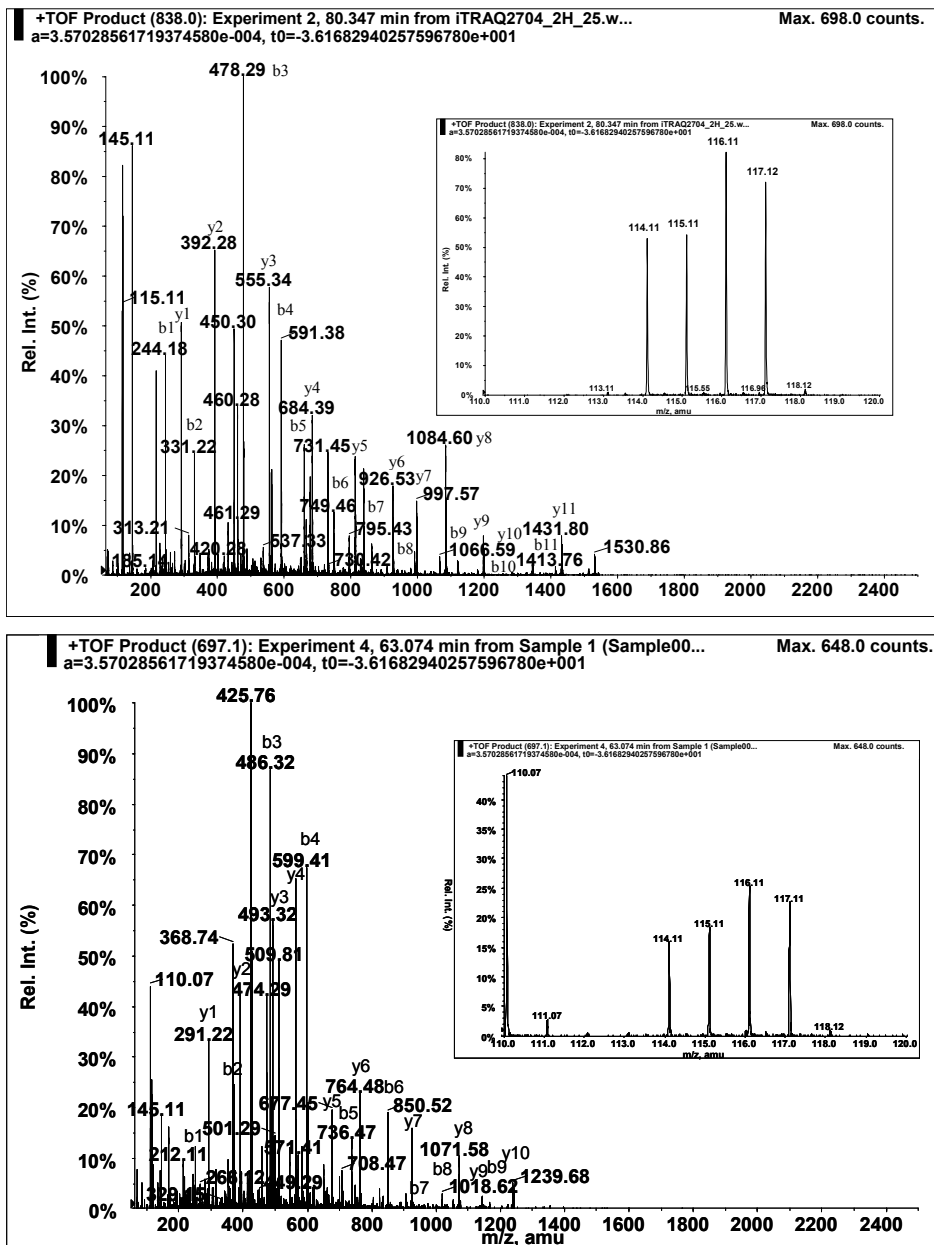


Figure S6. Examples of iTRAQ data for 2, 4, 6 and 8h clotting time is given. ApoA1 (a) and Complement C3 precursor (b) identified. In inserts 114, 115, 116 and 117 ions showed.

Mascot Paramaters; Database Swissprot,
Type of search : MS/MS Ion Search

Enzyme : None

Variable modifications : Q->p-E(N-term Q), E->p-E(N-term E), Oxidation(M), Phos (ST), Phos(Y)

Mass values : Monoisotopic

Protein Mass : Unrestricted

Peptide Mass Tolerance : ± 0.2 Da

Fragment Mass Tolerance: ± 0.2 Da

Max Missed Cleavages : 1

Instrument type : Default

Data File Name : D:\mascot mgf files\8hrsTCA.mgf Number of queries : 145

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Rank	Peptide
277	432.75	863.49	863.40	0.08	0	47	1	DFLAEGGGV
305	510.81	1019.60	1019.50	0.10	0	36	2	DFLAEGGGVR
307	525.79	1049.56	1049.47	0.10	0	23	1	EGDFLAEGGGV
310	539.32	1076.63	1076.52	0.10	0	47	1	GDFLAEGGGVR
314	554.30	1106.58	1106.49	0.09	0	28	1	GEGDFLAEGGGV
323	597.82	1193.63	1193.52	0.11	0	28	1	SGEGDFLAEGGGV
324	603.85	1205.68	1205.57	0.11	0	51	1	EGDFLAEGGGVR
340	655.34	1308.67	1308.55	0.12	0	24	1	DSGEGDFLAEGGGV
348	675.88	1349.74	1349.62	0.12	0	62	1	SGEGDFLAEGGGVR
366	730.85	1459.68	1459.55	0.13	0	20	3	ADSGEGDFLAEGGGV+Phos. (ST)
368	733.40	1464.78	1464.65	0.13	0	78	1	DSGEGDFLAEGGGVR
383	773.38	1544.75	1544.61	0.14	0	(72)	1	DSGEGDFLAEGGGVR+Phos. (ST)
389	808.90	1615.79	1615.65	0.14	0	74	1	ADSGEGDFLAEGGGVR+Phos. (ST)

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Rank	Peptide
311	546.27	1090.52	1090.42	0.10	0	18	2	QGVNDNEEGF+Q->p-E(N-term Q)
328	619.81	1237.61	1237.49	0.12	0	23	2	QGVNDNEEGFF+Q->p-E(N-term Q)
341	663.33	1324.64	1324.52	0.12	0	20	4	QGVNDNEEGFFS+Q->p-E(N-term Q)
357	698.85	1395.69	1395.56	0.14	0	23	1	QGVNDNEEGFFSA+Q->p-E(N-term Q)

Table 1. A list of identified peptides using TCA precipitation and LC-MS/MS by Mascot is shown. Top 2 from mascot hits.

1. FIBA HUMAN Mass: 94914 Total score: 1044 Peptides matched: 29,
(P02671) Fibrinogen alpha chain precursor [Contains: Fibrinopeptide A]

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Rank	Peptide
23	433.29	432.28	432.25	0.04	0	9	3	GGGV+iTRAQ4p(N-term)
307	447.30	892.58	892.48	0.10	0	42	1	FLAEGGGV+iTRAQ4p(N-term)
477	504.82	1007.62	1007.50	0.11	0	53	1	DFLAEGGGV+iTRAQ4p(N-term)
508	519.81	1037.60	1037.43	0.17	0	43	1	SGEGDFLAEGG
517	525.36	1048.70	1048.58	0.12	0	39	1	FLAEGGGVR+iTRAQ4p(N-term)
528	533.33	1064.65	1064.53	0.12	0	91	1	GDFLAEGGGV+iTRAQ4p(N-term)
551	544.81	1087.60	1087.53	0.07	0	35	1	TGKTFFGFFS
556	548.33	1094.64	1094.50	0.14	0	42	1	EGDFLAEGGG+iTRAQ4p(N-term)
594	582.89	1163.76	1163.61	0.15	0	34	1	DFLAEGGGVR+iTRAQ4p(N-term)
600	594.86	1187.70	1187.56	0.15	0	27	1	EGDFLAEGGGVR+E->p-E(N-term E)
608	597.88	1193.74	1193.57	0.17	0	63	1	EGDFLAEGGGV+iTRAQ4p(N-term)
624	611.40	1220.78	1220.63	0.16	0	32	1	GDFLAEGGGVR+iTRAQ4p(N-term)
659	620.86	1239.70	1239.54	0.16	0	59	1	DSGEGDFLAEG+iTRAQ4p(N-term)
667	626.39	1250.76	1250.59	0.17	0	68	1	GEGDFLAEGGGV+iTRAQ4p(N-term)
702	649.38	1296.74	1296.56	0.18	0	54	1	DSGEGDFLAEGG+iTRAQ4p(N-term)
727	669.91	1337.80	1337.62	0.18	0	88	1	SGEGDFLAEGGGV+iTRAQ4p(N-term)
738	675.93	1349.84	1349.62	0.22	0	82	1	SGEGDFLAEGGGVR
782	469.96	1406.85	1406.69	0.16	0	47	1	GEGDFLAEGGGVR+iTRAQ4p(N-term)
802	727.43	1452.85	1452.65	0.20	0	73	1	DSGEGDFLAEGGGV+iTRAQ4p(N-term)
808	730.88	1459.75	1459.55	0.20	0	57	2	ADSGEGDFLAEGGGV+Phos (ST)
825	498.97	1493.90	1493.72	0.18	0	(59)	1	SGEGDFLAEGGGVR+iTRAQ4p(N-term)
832	767.42	1532.83	1532.62	0.21	0	(20)	2	DSGEGDFLAEGGGV+iTRAQ4p(N-term); Phos (ST)
834	773.42	1544.83	1544.61	0.22	0	(37)	1	DSGEGDFLAEGGGVR+Phos (ST)
846	533.16	1596.47	1596.59	-0.12	0	8	9	GSSGTGGTATWKPSS+2 Phos(ST)
852	537.32	1608.95	1608.75	0.20	0	73	1	DSGEGDFLAEGGGVR+iTRAQ4p(N-term)
867	808.95	1615.88	1615.65	0.23	0	48	1	ADSGEGDFLAEGGGVR+Phos (ST)
918	561.01	1680.00	1679.79	0.21	0	(47)	1	ADSGEGDFLAEGGGVR+iTRAQ4p(N-term)
922	563.99	1688.94	1688.72	0.22	0	(32)	1	DSGEGDFLAEGGGVR+iTRAQ4p(N-term); Phos(ST)
960	587.67	1759.98	1759.75	0.23	0	(35)	1	ADSGEGDFLAEGGGVR+iTRAQ4p(N-term); Phos(ST)

2. FIBB HUMAN Mass: 55892 Total score: 466 Peptides matched: 13, (P02675)
Fibrinogen beta chain precursor [Contains: Fibrinopeptide B]

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Rank	Peptide
<u>551</u>	544.81	1087.60	1087.46	0.14	0	35	1	DNEEGFFS+iTRAQ4p (N-term)
<u>587</u>	580.33	1158.64	1158.50	0.15	0	24	1	DNEEGFFSA+iTRAQ4p (N-term)
<u>611</u>	601.83	1201.65	1201.50	0.15	0	43	1	NDNEEGFFS+iTRAQ4p (N-term)
<u>618</u>	607.86	1213.70	1213.54	0.16	0	46	1	VNDNEEGFF+iTRAQ4p (N-term)
<u>653</u>	619.83	1237.65	1237.49	0.16	0	43	1	QGVNDNEEGFF+Q->p-E (N-term Q)
<u>675</u>	636.37	1270.72	1270.56	0.16	0	53	1	GVNDNEEGFF+iTRAQ4p (N-term)
<u>678</u>	637.36	1272.71	1272.54	0.17	0	56	1	NDNEEGFFSA+iTRAQ4p (N-term)
<u>720</u>	663.35	1324.69	1324.52	0.16	0	58	1	QGVNDNEEGFFS+Q->p-E (N-term Q)
<u>743</u>	679.90	1357.78	1357.59	0.19	0	45	1	GVNDNEEGFFS+iTRAQ4p (N-term)
<u>751</u>	686.91	1371.80	1371.61	0.19	0	27	1	VNDNEEGFFSA+iTRAQ4p (N-term)
<u>765</u>	698.88	1395.75	1395.56	0.20	0	45	1	QGVNDNEEGFFSA+Q->p-E (N-term Q)
<u>767</u>	698.89	1395.77	1395.56	0.21	0	(30)	1	QGVNDNEEGFFSA+Q->p-E (N-term Q)
<u>787</u>	715.42	1428.82	1428.63	0.20	0	50	1	GVNDNEEGFFSA+iTRAQ4p (N-term)

Table 2. Peptides in the supernatant identified after precipitation and itraq labeling by LCMSMS and subsequent SwissProt database search by the MASCOT algorithm.

MASCOT parameters: Type of search: MS/MS Ion Search.

Enzyme : None; Variable modifications: Q->E(N-term Q), E->p-E(N-term E), iTRAQ4p(N-term), iTRAQ4p(Y), Oxidation(M), Phos(ST), Phos(Y); Mass values: Monoisotopic; Protein Mass: Unrestricted; Peptide Mass Tolerance: ± 0.3 Da; Fragment Mass Tolerance: ± 0.2 Da).

MASCOT Search Parameters

Type of search: MS/MS Ion Search

Enzyme: None

Variable modifications: Q->p-E (N-term Q), E->p-E (N-term E), iTRAQ4p(N-term), iTRAQ4p (Y), Oxidation (M), Phospho (ST), Phospho (Y)

Mass Monoisotopic

Protein Mass: Unrestricted

Peptide Mass Tolerance: ± 0.3 Da

Fragment Mass Tolerance: ± 0.2 Da

Max Missed Cleavages: 1

Instrument type: Default

Data File Name: D:\mascot mgf files\itraq_2_6hrs sup.mgf